Motivation

• Cornell Carbon Neutral Campus Initiative

• Reduce the need for parking

• Reduce congestion and other externalities associated with car traffic

• Reduce costs for commuter
My Research

• Determine the existing transit coverage
• Explore the sensitivity of the coverage to different parameters
• Propose better alternatives
  • New routes
  • New services (e.g. micro transit)
• Build the road network
• Build the bus network
• Merge the two
• Build simulation, modeling, and optimization tools
Generating the Road Network

• Tried a few different approaches
  • OSM
  • New York data repo
  • Settled on OSMnx

• Road network is represented as huge dictionary
  • G[node][neighbors][path][features]
  • Most important features
    • Road type
    • Length
    • Geometry
Getting Travel Times

• Each road has a type (motorway, primary, tertiary, residential, etc.)
  • Mapped types to speed limits if not already there
  • Used speed limit and distance to calculate time to traverse
• Need to add an edge transfer cost to simulate red lights and delays
• Technique worked well for shorter distance
  • <15 minutes was plus or minus 10% of google maps estimate
• For longer distances need a better solution for accurate times
• I believe Google Maps has some sort of API for this
The Road Network
Getting the bus network

• We obtained TCATs static GTFS feed
  • Google inspired data format for transit schedules
  • Contains schedules, stops, routes, route geometries, etc.

• Loaded this information into a database (there is a package for this)

• Constructed the routes for each bus
  • Routes alone weren’t sufficient for simulation

• Built a simulation for the bus system
  • Never actually used it
The Bus Network
Merging the Bus and Road Network

• We have geospatial data from both
• Must link the two
• Complication: (Blue = roads, yellow = bus stop)

Solution: Map matching
• If less than 20m from a node, map to node, else:
• Interpolate edge geometries between nearby nodes
• Select closest one
• Create node to represent bus stop
  • Updating road network was a challenge
Ithaca Merged Network
Analyzing Coverage

• What determines if an individual is covered?
  • Proximity
    • How close is the closest bus stop
  • Convenience
    • How much longer does it take to commute via bus than by car
  • Frequency
    • How often do busses come

*There are many other factors that contribute like weather or the need for child care but we lack the data to do this analysis.
Feasibility Constraints

• Distance
  • Must be a stop (or multiple) within the coverage radius of the origin

• Convenience
  • The time has to be less than the time it would take to commute to campus via car multiplied by a scaling factor
  • 15min; t<10min: t*2; t<20min: t*1.75; t>20min: t*1.5

• Frequency
  • Requires a certain number of bus trips from origin to Cornell between 7:00 and 10:00am
  • Requires a certain number of bus trips from Cornell to origin between 4:00 and 7:00pm
Creating Data

• For a while we lacked any actual data, so we created our own!
• I cut Tomkins county up into ~20 districts
• Googled the population of each, guessed the proportion of Cornell Faculty/staff based on proximity
• Created an algorithm to distribute the population based on the structure of the graph
  • Each node got a score based on
    • Number of incident edges
    • Length of incident edges
    • Other unique features
<table>
<thead>
<tr>
<th>Building</th>
<th>Class_Leve</th>
<th>Classifica</th>
<th>College_Co</th>
<th>Department</th>
<th>Facility_C</th>
<th>Match_addr</th>
<th>Rideshare</th>
<th>Subclass</th>
<th>geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>STATLER HOTEL</td>
<td>N/A</td>
<td>EMERITUS</td>
<td>N/A</td>
<td>SHA FIN ACC REALESTA</td>
<td>2033A</td>
<td>Rideshare</td>
<td>EMERITUS</td>
<td>POINT (-76.4163976305 32.4672580487)</td>
</tr>
<tr>
<td>1</td>
<td>WARREN HALL</td>
<td>N/A</td>
<td>EMPLOYEE</td>
<td>N/A</td>
<td>APPLIED ECONOMICS &amp;</td>
<td>1026</td>
<td>REGULAR/PAYROLL</td>
<td>POINT (-76.6931612007 39.8742.8595165267)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CORSON HALL</td>
<td>N/A</td>
<td>EMPLOYEE</td>
<td>N/A</td>
<td>ECOLOGY &amp; EVOLUTIONA</td>
<td>1019E</td>
<td>REGULAR/NOT ON PAYROLL</td>
<td>POINT (-76.5835345239 31.2342.5680684769)</td>
<td></td>
</tr>
</tbody>
</table>
Using Real data to determine coverage

- Mapped an individual to the nearest node
- Determined all the bus stops within the coverage radius
- Determined all busses that stop at any stop in CR and Cornell
  - Determine specific trips that run from node in CR to Cornell stop in morning
    - Check if the trip time meets the convenience constraint
      - Add trip to feasible trips
  - Do the same from Cornell to stop in CR in afternoon.
TODO: Integrate new data

- We rely on balance between not counting walking or parking times
- We received new data with coordinates for every building and parking lot
  - We hope to also receive data for which parking lots are permitted under which permit
- Next step is to integrate where someone works and where they are allowed to park into the calculation
  - Penalty for walking to and from bus stop
  - Require route to stop within radius of destination
  - Penalty for parking lot to building
Loosest Constraints

- 1200m coverage radius
- 1 stop in morning
- 1 stop in afternoon

*80% coverage of Cornell faculty and staff
Tight Constraints

- 500m coverage radius
- Every 30 minutes
- 15min; t<10min: t*2; t<20min: t*1.75; t>20min: t*1.5
## Results

<table>
<thead>
<tr>
<th>Coverage Radius/Frequency</th>
<th>250 meters</th>
<th>500 meters</th>
<th>750 meters</th>
<th>1200 meters (¾ mile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every 30 min</td>
<td>13.6%</td>
<td>31.6%</td>
<td>42.0%</td>
<td>50.4%</td>
</tr>
<tr>
<td>Every 20 min</td>
<td>10.3%</td>
<td>26.0%</td>
<td>33.4%</td>
<td>43.9%</td>
</tr>
<tr>
<td>Every 15 min</td>
<td>8.4%</td>
<td>20.0%</td>
<td>30.4%</td>
<td>42.8%</td>
</tr>
</tbody>
</table>

*Only Cornell Faculty and Staff

**Household locations are mapped to nodes
Predictive modeling

• Using Random Forest on the following features
  • Proximity to closest bus stop
  • Proximity to second closest bus stop
  • Number of bus stops in 1km radius
  • Average afternoon trip duration (via bus)
  • Average morning trip duration (via bus)
  • Total number of feasible trips (Min between morning and afternoon)
  • Travel time (via car)
  • The ratio between bus and car travel time
  • The department someone is in
  • Faculty/staff status
Actual Distribution

Green: No parking pass
Red: parking pass
Actual Distribution

Green: No parking pass
Red: parking pass
Predicted Distribution

Green: No parking pass
Red: parking pass
Predicted Distribution

Green: No parking pass
Red: parking pass
Correctness

Green: Correct
Red: Predicted to have parking pass when didn’t
Blue: Predicted to not have parking pass but does
Correctness

Green: Correct
Red: Predicted to have parking pass when didn’t
Blue: Predicted to not have parking pass but does
Model Comparison

<table>
<thead>
<tr>
<th>Predicted Holder</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Holder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>42</td>
<td>533</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>1227</td>
</tr>
</tbody>
</table>

**SVM**
- 69-70% accuracy

<table>
<thead>
<tr>
<th>Predicted Holder</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Holder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>184</td>
<td>391</td>
</tr>
<tr>
<td>1</td>
<td>79</td>
<td>1185</td>
</tr>
</tbody>
</table>

**Random Forest**
- 76-78% accuracy

*Baseline is 69.5%*
Model Analysis

• Feature importance varied widely when hyperparameters were changed

• Most important: Total number of feasible trips

• Important
  • Travel time (car)
  • Number of bus stops
  • Department*

• Not Important
  • Proximity to closest/second closest
  • Average travel time (bus)
  • Travel ratio
  • Employee status
Next Steps

• Integrate new data to improve model of feasibility
• Use this to improve predictive modeling
• Develop an optimization framework to add best routes